Zero-point quantum fluctuations and dark energy

Michele Maggiore

Département de Physique Théorique, Université de Genève, CH-1211 Geneva, Switzerland

In the Hamiltonian formulation of General Relativity the energy associated to an asymptotically flat space-time with metric $g_{\mu\nu}$ is related to the Hamiltonian $H_{\rm GR}$ by $E = H_{\rm GR}[g_{\mu\nu}] - H_{\rm GR}[\eta_{\mu\nu}]$, where the subtraction of the flat-space contribution is necessary to get rid of an otherwise divergent boundary term. This classic result indicates that the energy associated to flat space does not gravitate. We apply the same principle to study the effect of zero-point fluctuations of quantum fields in cosmology, proposing that their contribution to the cosmic expansion is obtained computing the vacuum energy of quantum fields in a FRW space-time with Hubble parameter $H(t)$ and subtracting from it the flat-space contribution. Then the term proportional to Λ_c^4 (where Λ_c is the UV cutoff) cancels and the remaining (bare) value of the vacuum energy density is proportional to $\Lambda_c^2 H^2(t)$. After renormalization, this produces a renormalized vacuum energy density $\sim M^2 H^2(t)$, where M is the scale where quantum gravity sets is, so for M of order of the Planck mass a vacuum energy density of the order of the critical density can be obtained without any fine tuning. The counterterms can be chosen so that the renormalized energy density and pressure satisfy $p = w\rho$, with w a parameter that can be fixed by comparison to the observed value, so in particular one can chose $w = -1$. An energy density evolving in time as $H^2(t)$ is however observationally excluded as an explanation for the dominant dark energy component which is responsible for the observed acceleration of the universe. We rather propose that zero-point vacuum fluctuations provide a new subdominant "dark" contribution to the cosmic expansion that, for a UV scale M slightly smaller than the Planck mass, is consistent with existing limits and potentially detectable.

I.   INTRODUCTION

The cosmological constant problem is a basic issue of modern cosmology (see e.g. [1–4]). One aspect of the problem is to understand what is the effect of the zero-point fluctuations of quantum fields on the cosmic expansion. Zero-point quantum fluctuations seem to give a contribution to the vacuum energy density of order Λ_c^4, where Λ_c is the UV cutoff. Even for a cutoff as low as $\Lambda_c \sim 1\,{\rm TeV}$, corresponding to scales where quantum field theory is well tested, this exceeds by many orders of magnitude the value of the critical density of the universe, which is of order $(10^{-3}\,{\rm eV})^4$. This aspect of the problem has by now a long history, which even goes back to works of Nernst and of Pauli in the 1920s [2].

It is clear that this Λ_c^4 contribution should somehow be subtracted. Renormalization in quantum field theory (QFT) offers a logically viable possibility. One should not forget, in fact, that this cutoff-dependent contribution, as any similar quantity in QFT, is just a bare quantity rather than a physical observable. Working for instance with a cutoff Λ_c in momentum space one finds that, in Minkowski spacetime, the bare, cut-off dependent, contribution to the vacuum energy density is

$$\rho_{\rm bare}(\Lambda_c) = c\Lambda_c^4\,, \qquad (1)$$

with c some constant that depends on the number and type of fields of the theory. The standard procedure in QFT is to add to it a cutoff-dependent counterterm, chosen so that it cancels the divergent part and leaves us with the desired finite part (for an elementary discussion of this point in the cosmological constant case, see e.g. the textbook [5]), i.e.

$$\rho_{\rm count}(\Lambda_c) = -c\Lambda_c^4 + \rho_{\rm vac}\,, \qquad (2)$$

where $\rho_{\rm vac} \sim (10^{-3}{\rm eV})^4$ is independent of the cutoff and equal to the observed value of the vacuum energy density (assuming that vacuum energy is indeed responsible for the observed acceleration of the universe). The physical, renormalized, vacuum energy density

$$\rho_{\rm ren} \equiv \rho_{\rm bare}(\Lambda_c) + \rho_{\rm count}(\Lambda_c) \qquad (3)$$

is therefore by definition independent of the cutoff and equal to the observed value $\rho_{\rm vac}$. In principle this procedure is not different from what is done when one renormalizes, say, the electron mass or the electron charge in QED, where again with suitable counterterms one removes the divergent, cut-off dependent, parts and fixes the finite parts so that they agree with the observed values (in this sense the statement, often heard, that QFT gives a *wrong* prediction for the cosmological constant, is not correct. Strictly speaking QFT makes no prediction for the cosmological constant, just as it does not predict the electron mass nor the fine structure constant).

Still, for the vacuum energy this procedure is not really satisfying. The source of uneasiness is partly due to the fact that the counterterm must be fine tuned to a huge precision in order to cancel the Λ_c^4 term and leave us with a much smaller result. In fact, if in eq. (1) we take Λ_c at least larger than a few TeV, where quantum field theory has been successfully tested, $\rho_{\rm bare}$ is at least of order $({\rm TeV})^4 = (10^{12}\,{\rm eV})^4$, so $\rho_{\rm count}$ in eq. (3) must be fine tuned so that it cancels something of order $(10^{12}\,{\rm eV})^4$, leaving a result of order $(10^{-3}\,{\rm eV})^4$. This fine tuning becomes even much worse if one dares to take Λ_c of the order of the Planck mass $M_{\rm Pl} \sim 10^{19}$ GeV. The fact that the counterterm must be tuned to such a precision creates a naturalness problem. Here, however, one might argue that neither the bare term nor the counterterm have any

physical meaning and only their sum is physical, so this fine tuning is different from an unplausible cancellation between observable quantities. The same kind of cancellation appears, for instance, when one computes the Casimir effect. The crucial point, however, is that this renormalization procedure leaves us with no clue as to the physical value that emerges from this cancellation, so it gives no explanation of why the energy density associated to the cosmological constant appears to have just a value of the order of the critical density of the universe at the present epoch.

The point of view that we develop in this paper is that, even if renormalization must be an ingredient for understanding the physical effects of vacuum fluctuations, it is not the end of the story. The Casimir effect mentioned above gives indeed a first hint of what could be the missing ingredient for a correct treatment of vacuum energies in cosmology. In the Casimir effect the quantity that gives rise to observable effects, which have indeed been detected experimentally, is the difference between the vacuum energy in a given geometry (e.g., for the electromagnetic field, between two parallel conducting plates) and the vacuum energy computed in a reference geometry, which is just flat space-time in an infinite volume. Both terms are separately divergent as Λ_c^4, but their difference is finite and observable. This can suggest that, to obtain the physical effect of the vacuum energy density on the expansion of the universe, one should similarly compute the vacuum energy density in a FRW space-time and subtract from it the value of a reference space-time, which is naturally taken as Minkowski space.

A possible objection to this procedure could be that General Relativity requires that any form of energy should be a source for the gravitational field, which seems to imply that even the vacuum energy associated to flat space should contribute. A more careful look at the formalism of GR shows however that the issue is not so simple and that, in a sense, the subtraction that we are advocating is in fact already part of the standard tenets of classical GR. To define carefully the energy associated to a field configuration in GR, it is convenient to use the Hamiltonian formulation, which goes back to the classic paper by Arnowitt, Deser and Misner [6] (ADM). As we will review in more detail in Section II, in order to define the Hamiltonian of GR one must at first work in a finite three-dimensional volume, and then the Hamiltonian takes the form

$$H_{\rm GR} = H_{\rm bulk} + H_{\rm boundary} \,, \qquad (4)$$

where $H_{\rm bulk}$ is given by an integral over the three-dimensional finite spatial volume at fixed time, and $H_{\rm boundary}$ by an integral over its (two-dimensional) boundary. At this point one would like to define the energy of a classical field configuration as the value of this Hamiltonian, evaluated on the classical solution, but one finds both a surprise and a difficulty. The "surprise" (which actually is just a consequence of the invariance under diffeomorphisms) is that $H_{\rm bulk}$, evaluated on any

classical solution of the equations of motion, vanishes, so the whole contribution comes from the boundary term. The difficulty is that the boundary term, evaluated on any asymptotically flat metric (including flat space-time) diverges when the boundary is taken to infinity. The solution proposed by ADM is to subtract from this boundary term, evaluated on an asymptotically flat space-time with metric $g_{\mu\nu}$, the same term computed in flat space-time. The corresponding energy (or mass) is finite and is known as the ADM mass, and provides the standard definition of mass in GR. For instance, when applied to the Schwarzschild space-time, the ADM mass computed in this way turns out to be equal to the mass M that appears in the Schwarzschild metric.

The ADM prescription can be summarized by saying that, in GR, the energy E associated to an asymptotically flat space-time with metric $g_{\mu\nu}$ can be obtained from the Hamiltonian $H_{\rm GR}$ by

$$E = H_{\rm GR}[g_{\mu\nu}] - H_{\rm GR}[\eta_{\mu\nu}] \,, \qquad (5)$$

where $\eta_{\mu\nu}$ is the flat metric. Even if the context in which this formula is valid, namely asymptotically flat space-times, is different from the cosmological context in which we are interested here, still eq. (5) suffices to make the point that the intuitive idea that GR requires that any form of energy acts as a source for the gravitational field is not really correct. Equation (5) tells us that the energy associated to Minkowski space does not gravitate.

The idea of this paper is to generalize eq. (5) to the case of zero-point quantum fluctuation in curved space, proposing that the effect of zero-point fluctuations on the cosmic expansion should be obtained by computing the vacuum energy of quantum fields in a FRW space-time with Hubble parameter $H(t)$ and subtracting from it the flat-space contribution. Computed in this way, the physical effect of zero-point fluctuations on the cosmic expansion can be seen as a sort of "cosmological Casimir effect": while in the standard Casimir effect one computes the vacuum energy in a given geometry (say, for the electromagnetic field, between two infinite parallel conducting planes) and subtracts from it the value computed in a reference geometry (flat space-time in an infinite volume), here we compute the vacuum energies of the various fields in a given curved space-time, e.g. in FRW space-time, and we subtract from it the value computed in a reference space-time, i.e. Minkowski.

A possible objection to the idea of applying eq. (5) to vacuum fluctuations could be that one might think that, after all, the structure of UV divergences is determined by local properties of the theory, while whether a space-time is Minkowski is a global question. However one should keep in mind that, even if $T_{\mu\nu}(x)$ is a local quantity, the vacuum expectation value $\langle 0|T_{\mu\nu}|0\rangle$ in a curved background involves the global aspects of the space-time in which it is computed. This is due to the fact that $\langle 0|T_{\mu\nu}|0\rangle$ requires a definition of the vacuum state $|0\rangle$. The vacuum is defined from the condition $a_{\bf k}|0\rangle = 0$, where the annihilation operators $a_{\bf k}$ are defined with re-

spect to a set of mode functions $\phi_{\mathbf{k}}(t)$. The mode functions, in turn, are obtained solving a wave equation over the whole space-time, and therefore are sensitive to global properties of the space-time itself. In particular, in a FRW background the mode functions depend on the scale factor so their time derivatives (which enter in the computation of $\langle 0|T_{\mu\nu}|0\rangle$) depend on the Hubble parameter $H(t)$. As we will review below, this results in the fact that the quadratic divergence in $\langle 0|T_{\mu\nu}|0\rangle$ depends on the expansion rate $H(t)$ of the FRW background.

The paper is organized as follows. In Section II we recall how eq. (5) is derived in the ADM formalism. The reader familiar with the subject, or not interested in the derivation, might simply wish to move directly to Section III, where we apply this classical subtraction, together with standard renormalization theory, to zero-point vacuum fluctuations. Some cosmological consequences of our proposal are discussed in Section IV, while Section V contains our conclusions.

II. SUBTRACTIONS IN CLASSICAL GR: THE ADM MASS

In this section we briefly discuss how the Arnowitt-Deser-Misner (ADM) mass is defined in GR [6] (we follow the very clear discussion of the textbook [7]). We begin by recalling that, in a finite four-dimensional volume $\mathcal{V}$ with boundary $\partial\mathcal{V}$, the gravitational action is (setting $c = 1$ and using the signature $\eta_{\mu\nu} = (-, +, +, +)$)

$$S_{\text{grav}} = \frac{1}{16\pi G}\int_{\mathcal{V}} d^4x\,\sqrt{-g}\,R + \frac{1}{8\pi G}\int_{\partial\mathcal{V}} d^3y\,\sqrt{|h|}\epsilon K\,,$$
(6)

where $g_{\mu\nu}$ is the four-dimensional metric, h_{ij} is the metric induced on the boundary $\partial\mathcal{V}$, $h = \det h_{ij}$, K is the trace of the extrinsic curvature of the boundary, y^i are the coordinates of the boundary, and $\epsilon = +1$ on the regions of the boundary where $\partial\mathcal{V}$ is time-like and $\epsilon = -1$ where $\partial\mathcal{V}$ is space-like. The first term is the usual Einstein-Hilbert action, while the second is a boundary term which is necessary to obtain a well-defined variational principle.

To pass to the Hamiltonian formalism one performs the $3 + 1$ decomposition of the metric,

$$ds^2 = -\alpha^2 dt^2 + h_{ij}(dx^i + \beta^i dt)(dx^j + \beta^j dt)\,,$$
(7)

where α and β^i are the lapse function and shift vector, respectively, and h_{ij} is the induced metric on the 3-dimensional spatial hypersurfaces. One defines as usual the conjugate momentum $\pi^{ij} = \partial\mathcal{L}/\partial\dot{h}_{ij}$, where $\mathcal{L}$ is the "volume part" of the Lagrangian density, and the Hamiltonian H is then the volume integral of the Hamiltonian density $\mathcal{H} = \pi^{ij}\dot{h}_{ij} - \mathcal{L}$. The explicit computation (see e.g. Section 4.2.6 of [7]) gives

$$(16\pi G)\,H = \int_{\Sigma_t} d^3x\sqrt{h}\left(-\alpha C_0 - 2\beta_i C^i\right)$$
(8)
$$- 2\int_{S_t} d^2\theta\sqrt{\sigma}\left[\alpha k - \beta_i r_j(K^{ij} - Kh^{ij})\right]\,,$$

where Σ_t denotes the three-dimensional spatial hypersurfaces at fixed t and S_t (with coordinates θ^i) is the intersection of Σ_t with $\partial\mathcal{V}$. In an asymptotically flat space-time S_t is just a 2-sphere at large radius $r = R$ and fixed t; C_0 and C^i depend only on h_{ij} and on its derivatives, but not on α and β_i, while k in eq. (8) is the trace of the extrinsic curvature of S_t, and σ is the determinant of the two-dimensional induced metric.

One would like to define the energy of a classical field configuration, i.e. of a solution of the equations of motion, as the value of this Hamiltonian on the solution. Performing the variation with respect to α and β_i one obtains the constraint equations $C_0 = 0$ and $C^i = 0$. Therefore, on a classical solution, the volume term in eq. (8) vanishes, and only the boundary term contributes. The ADM mass is defined by setting $\alpha = 1$ and $\beta_i = 0$ after performing the variation (corresponding to the fact that energy is associated to asymptotic time translations; setting $\alpha = 0$ and $\beta^i = 1$ one rather obtains the ADM momentum P^i_{ADM}), so only the term proportional to k contributes to the mass. However, even for Minkowski space, this boundary term diverges. In fact, for an asymptotically flat space-time we can take $\partial\mathcal{V}$ to be a three-dimensional cylinder made of the two three-dimensional time-like hypersurfaces $\{t = t_1, r \leq R\}$ and $\{t = t_2, r \leq R\}$ (the "faces" of the three-dimensional cylinder) and of the space-like hypersurface $\{r = R, t_1 \leq t \leq t_2\}$. Let us denote by K_0 the extrinsic curvature of $\partial\mathcal{V}$ computed with a flat Minkowski metric. The faces at $t = t_1$ and t_2 have $K_0 = 0$; however, on the surface $\{r = R, t_1 \leq t \leq t_2\}$ the extrinsic curvature is $K_0 = 2/R$, while $|h|^{1/2} = R^2 \sin^2\theta$ and $\epsilon = +1$, so the boundary term in eq. (7) is [7]

$$\int_{\partial\mathcal{V}} d^3y\,\sqrt{|h|}\epsilon K_0 = 8\pi(t_2 - t_1)R\,,$$
(9)

and diverges both when $R \to \infty$ and when $(t_2 - t_1) \to \infty$. As a consequence, also the boundary term proportional to k in eq. (8) diverges, already for the Minkowski metric, and then of course also for generic asymptotically flat space-times. The ADM prescription is then to replace K in eq. (6) by $(K - K_0)$, i.e. to subtract from the trace of the extrinsic curvature computed with the desired metric, the value computed in flat space. Correspondingly, the trace k of the two-dimensional extrinsic curvature in eq. (8) becomes $(k - k_0)$, and the ADM mass of an asymptotically flat space-time is defined as [6]

$$M_{\text{ADM}} = -\frac{1}{8\pi G}\lim_{S_t \to \infty}\int_{S_t} d^2\theta\sqrt{\sigma}(k - k_0)\,.$$
(10)

If for instance one applies this definition to the Schwarzschild space-time, one finds that M_{ADM} is equal to the mass M which appears in the Schwarzschild metric.

What we learn from this is that, in classical GR, the mass or the energy that acts as the source of curvature of space-time, such as for instance the mass M that enters in the Schwarzschild metric, can be obtained from a

Hamiltonian treatment only after subtracting a flat-space contribution that need not be zero, and is in fact even divergent, but still does not act as a source of curvature.

III. APPLICATION TO ZERO-POINT ENERGIES

It is natural to apply the same principle to zero-point quantum fluctuations. In particular, in a cosmological setting, we propose that the zero-point energy density and pressure that contribute to the cosmological expansion are obtained by computing the energy density and pressure due to zero-point quantum fluctuations in a FRW metric with Hubble parameter $H(t) = \dot{a}/a$, and subtracting from it the flat-space contribution.

We consider first the contribution of a real massless scalar field. In a FRW background the mode expansion of the field is

$$\phi(x) = \int \frac{d^3 k}{(2\pi)^3 \sqrt{2k}} \left[a_{\mathbf{k}} \phi_k(t) e^{i\mathbf{k}\cdot\mathbf{x}} + a_{\mathbf{k}}^\dagger \phi_k^*(t) e^{-i\mathbf{k}\cdot\mathbf{x}} \right] , \tag{11}$$

where $\mathbf{k}$ is the comoving momentum, and $\phi_k(t)$ satisfies the massless Klein-Gordon equation in a FRW background,

$$\phi_k'' + 2\frac{a'}{a}\phi_k' + k^2 \phi_k = 0 , \tag{12}$$

where the prime is the derivative with respect to conformal time η. Writing $\phi_k = \psi_k/a$ this equation is reduced to the form

$$\psi_k'' + \left(k^2 - \frac{a''}{a} \right) \psi_k = 0 . \tag{13}$$

In a De Sitter background we have $a(\eta) = -1/(H\eta)$ so $a''/a = 2/\eta^2$, while in a matter-dominated (MD) epoch $a \sim \eta^2$ and therefore again $a''/a = 2/\eta^2$. Thus, in both cases the positive frequency solution of eq. (13) is

$$\psi_k(\eta) = \left(1 - \frac{i}{k\eta} \right) e^{-ik\eta} , \tag{14}$$

and therefore $\phi_k(\eta)$ is given by

$$\phi_k(\eta) = \frac{1}{a(\eta)} \left(1 - \frac{i}{k\eta} \right) e^{-ik\eta} , \tag{15}$$

with $a(\eta)$ the scale factor of De Sitter or MD epoch, respectively. In contrast, during a radiation-dominated (RD) epoch $a \sim \eta$, so $a''/a = 0$. Then $\psi_k(\eta) = e^{-ik\eta}$ and

$$\phi_k(\eta) = \frac{1}{a(\eta)} e^{-ik\eta} . \tag{16}$$

Using the energy–momentum tensor of a minimally coupled massless scalar field,

$$T_{\mu\nu} = \partial_\mu \phi \partial_\nu \phi - \frac{1}{2} g_{\mu\nu} g^{\rho\sigma} \partial_\rho \phi \partial_\sigma \phi , \tag{17}$$

and setting $g_{\mu\nu} = (-1, a^2 \delta_{ij})$, a simple computation [8, 9], reviewed in Appendix A, shows that the off-diagonal elements of $\langle 0|T_{\mu\nu}|0\rangle$ vanish, while the vacuum energy density and pressure are given by

$$\rho = \frac{1}{2} \int \frac{d^3 k}{(2\pi)^3 2k} \left(|\dot{\phi}_k|^2 + \frac{k^2}{a^2} |\phi_k|^2 \right) , \tag{18}$$

$$p = \frac{1}{2} \int \frac{d^3 k}{(2\pi)^3 2k} \left(|\dot{\phi}_k|^2 - \frac{k^2}{3a^2} |\phi_k|^2 \right) , \tag{19}$$

where the dot denotes the derivative with respect to cosmic time t. As they stand, these expressions must still be regularized, and we regularize them by putting a cutoff in momentum space. Recall that the comoving momentum k is just a label of the Fourier mode under consideration, while the physical momentum of the mode is given by k/a. We expect that quantum gravity enters the game when the *physical* momenta exceed the Planck scale, and we therefore put a time-independent cutoff Λ_c over physical, rather than comoving, momenta. In terms of comoving momentum k this means $k < a(t)\Lambda_c$. Since the modes ϕ_k depend only on $k = |\mathbf{k}|$, the angular integrals are trivially performed, and we finally get

$$\rho = \frac{1}{8\pi^2} \int_0^{a\Lambda_c} dk \, k \left(|\dot{\phi}_k|^2 + \frac{k^2}{a^2} |\phi_k|^2 \right) , \tag{20}$$

$$p = \frac{1}{8\pi^2} \int_0^{a\Lambda_c} dk \, k \left(|\dot{\phi}_k|^2 - \frac{k^2}{3a^2} |\phi_k|^2 \right) . \tag{21}$$

A. Vacuum fluctuations in De Sitter space

1. Vacuum energy density

We compute first these expressions in De Sitter space. We therefore plug eq. (15), with $a(\eta) = -1/(H\eta)$, into eqs. (20) and (21). For the energy density the result is

$$\begin{aligned} \rho(\Lambda_c) &= \frac{1}{4\pi^2} \int_0^{a\Lambda_c} dk \, k \left(\frac{k^2}{a^4} + \frac{H^2}{2a^2} \right) \\ &= \frac{\Lambda_c^4}{16\pi^2} + \frac{H^2 \Lambda_c^2}{16\pi^2} . \end{aligned} \tag{22}$$

The first term is the well known flat-space result, proportional to the fourth power of the cutoff, while the term quadratic in Λ_c is the correction due to the expansion of the universe. The appearance of a term $\sim H^2 \Lambda_c^2$, in a theory with two scales, the UV cutoff Λ_c and the Hubble scale H, can also be understood using rather general arguments [10].

According to our proposal we now subtract the flat-space contribution, which is simply the term $\sim \Lambda_c^4$ in eq. (22), and we find that in De Sitter space the zero-point quantum fluctuations of a real scalar field have a (bare) energy density

$$\rho_{\text{bare}}(\Lambda_c) = \frac{H^2 \Lambda_c^2}{16\pi^2} . \tag{23}$$

5

The subscript "bare" stresses that this is still a bare, cut-off dependent quantity. We have eliminated the quartic divergence thanks to the "classical" prescription (5), but the result is still a bare energy density, diverging quadratically with the cutoff. In this sense, the situation is different from the usual Casimir effect, where the subtraction of the flat-space contribution suffices to make the result finite. However, we can now use standard renormalization theory, so the renormalized energy density is obtained by adding a counterterm whose divergent part is chosen so to cancel this quadratic divergence, and whose finite part $\rho_{\rm vac}$ is in principle fixed by the observation, so

$$\rho_{\rm count}(\Lambda_c) = -\frac{H^2\Lambda_c^2}{16\pi^2} + \rho_{\rm vac}\,. \qquad (24)$$

As usual in the renormalization procedure, the finite part $\rho_{\rm vac}$ cannot be predicted. It must be fixed to the observed value. What we have gained, with respect to eq. (2), is that now we have a different understanding of what is a "natural" value of this finite part. If quantum gravity sets in at a mass scale M (so that M could be typically given by the Planck mass $M_{\rm Pl}$, or by the string mass), a natural, non fine-tuned value of $\rho_{\rm vac}$ is given by

$$\rho_{\rm vac} = {\rm const.} \times \sigma\frac{H^2M^2}{16\pi^2}\,, \qquad (25)$$

where "const." is a constant $\mathcal{O}(1)$, which cannot be fixed by naturalness arguments only, and must be determined by comparison with the experiment.[41]

Observe that naturalness arguments cannot fix the sign of the finite part either, and we have added a factor $\sigma = \pm 1$ to take this fact explicitly into account. There is in fact no a priori reason why the renormalized vacuum energy should necessarily be positive. For instance, the vacuum energies obtained for fields in a finite volume from the Casimir effect can be either positive or negative, depending e.g. on the type of field and on the geometry considered.

2. Equation of state and general covariance

Repeating the same computation for the pressure we get

$$\begin{aligned} p(\Lambda_c) &= \frac{1}{4\pi^2}\int_0^{a\Lambda_c} dk\,k\left(\frac{k^2}{3a^4} - \frac{H^2}{6a^2}\right) \\ &= \frac{\Lambda_c^4}{48\pi^2} - \frac{H^2\Lambda_c^2}{48\pi^2}\,, \end{aligned} \qquad (26)$$

The term $\sim \Lambda_c^4$ in eq. (26) was already computed in [11] and is just the result in Minkowski space, that we subtract, so we end up with

$$p_{\rm bare}(\Lambda_c) = -\frac{H^2\Lambda_c^2}{48\pi^2}\,. \qquad (27)$$

Observe that $p_{\rm bare}(\Lambda_c) = -(1/3)\rho_{\rm bare}(\Lambda_c)$. However, it would be incorrect to conclude that vacuum fluctuations

in De Sitter space satisfy the equation of state $p = w\rho$ with $w = -1/3$. The point is that this relation only holds for the bare quantities, and not necessarily for the renormalized ones. The physical, renormalized, pressure is obtained by adding a counterterm

$$p_{\rm count}(\Lambda_c) = +\frac{H^2\Lambda_c^2}{48\pi^2} + p_{\rm vac}\,. \qquad (28)$$

Observe that regularizing the theory with a cutoff over spatial momenta, as we have done, breaks explicitly Lorentz invariance in Minkowski space, since the notion of maximum value of spatial momenta is not invariant under boosts. In a generic FRW background, of course, Lorentz transformations are not a symmetry of the theory, since the metric depends explicitly on time, and the guiding principle is rather general covariance, which again is broken by a cutoff over spatial momenta. However, even if the regularization breaks general covariance, in De Sitter space a generally covariant result can still be obtained in the end for the physical, renormalized, vacuum expectation value of the energy-momentum tensor, just by choosing the finite parts of the counterterms such that $p_{\rm vac} = -\rho_{\rm vac}$. Then the vacuum expectation value of the renormalized energy-momentum tensor $T^\mu_\nu = {\rm diag}(-\rho, p, p, p)$ becomes

$$\langle 0|T^\mu_\nu|0\rangle = {\rm const.} \times \sigma\frac{H^2M^2}{16\pi^2}(-\delta^\mu_\nu)\,, \qquad (29)$$

or, lowering the upper index with the metric $g_{\mu\nu}$,

$$\langle 0|T_{\mu\nu}|0\rangle = {\rm const.} \times \sigma\frac{H^2M^2}{16\pi^2}(-g_{\mu\nu})\,. \qquad (30)$$

Since in De Sitter space H is constant we see that, with the choice $p_{\rm vac} = -\rho_{\rm vac}$, $\langle 0|T_{\mu\nu}|0\rangle$ is just given by a numerical constant times $g_{\mu\nu}$, and it is therefore covariantly conserved. In the language of the effective action for gravity, which is obtained by treating the metric $g_{\mu\nu}$ as a classical background and integrating over the matter degrees of freedom (see e.g. refs. [12–14]), the vacuum expectation value of the energy-momentum tensor is given by the functional derivative $(2/\sqrt{-g})\delta/\delta g_{\mu\nu}$ of the effective action. Then a contribution such as that given in eq. (30) can be obtained by taking the functional derivative of the volume term in the effective action (see ref. [15] for a discussion of the equation of state that can be obtained from the various contributions to the effective action).

The choice $p_{\rm vac} = -\rho_{\rm vac}$ will therefore be assumed in the following, for De Sitter space-time. Observe that a covariant result for the renormalized value of $\langle 0|T_{\mu\nu}|0\rangle$ is obtained with a counterterm that is not covariant, i.e. is not proportional to $g_{\mu\nu}$, since $p_{\rm count}(\Lambda_c) \neq -\rho_{\rm count}(\Lambda_c)$, but again this is a consequence of the fact that our regularization is not covariant.

3. Contribution of higher-spin fields and massive particles

A similar conclusion about the "natural" value of the energy density of vacuum fluctuations holds if we add the contribution of fields with different spin. In fact, massless fermions and gauge bosons have a conformally invariant action and then, since the FRW metric is conformally equivalent to flat Minkoswki space, their vacuum energy in a FRW space-time is the same as in Minkowski space. Therefore, with our subtraction (5), they do not contribute to $\rho_{\rm vac}$ and $p_{\rm vac}$. This is completely equivalent to the well known fact that vacuum fluctuations of massless fermions and gauge bosons are not amplified during inflation.

The contribution of massive fermions to vacuum fluctuations in De Sitter space, as well as the generalization of eq. (22) to massive bosons, can be readily computed. For massive bosons, eq. (22) becomes [9, 16]

$$\rho_B(\Lambda_c) = \frac{1}{4\pi^2} \int_0^{a\Lambda_c} dk\, k \left[\frac{k\omega_k}{a^3} + \frac{H^2 k}{2a^3 \omega_k} \left(1 + \frac{m^2}{\omega_k^2} \right) \right],$$
(31)

where we have neglected terms of order H^4, and terms convergent in the UV; ω_k is defined as

$$\omega_k = \sqrt{m^2 + (k/a)^2}\,,$$
(32)

so in the massless limit $\omega_k \to k/a$ and we recover eq. (22) [42]. In this massive case our prescription amounts to subtracting the flat-space contribution given by the term $k\omega_k/a^3$ in brackets. The remaining, quadratically divergent term, leads again to eq. (25), times a correction $[1 + \mathcal{O}(m^2/M^2)]$, and to a logarithmically divergent term, which after renormalization produces a contribution to $\rho_{\rm vac}$ proportional to $H^2 m^2 \log(M/m)$, and therefore subleading for $m \ll M$.

The contribution of a massive Majorana spinor field to the vacuum energy density in De Sitter space is instead [16]

$$\rho_F(\Lambda_c) = \frac{1}{2\pi^2} \int_0^{a\Lambda_c} dk\, k \left[-\frac{k\omega_k}{a^3} + \frac{m^2 H^2 k}{8a^3 \omega_k^3} \right].$$
(33)

The first term in brackets gives the usual negative contribution to vacuum energy in flat space due to fermions. Observe that in a supersymmetric model, where to each Majorana spinor is associated a complex scalar field, and therefore two real scalar fields, the contribution $(-k\omega_k/a^3)$ in eq. (33) cancels exactly the bosonic contribution $(+k\omega_k/a^3)$ in eq. (31), giving the usual cancellation of vacuum energy for a supersymmetric theory in Minkowski space. In a realistic model with supersymmetry broken at a scale $\Lambda_{\rm susy}$, this cancellation is however only partial and leaves the usual result of order $\Lambda_{\rm susy}^4$. In our approach, instead, the quartic divergence are eliminated exactly by the prescription of subtracting the flat-space contribution.

The term that remains in eq. (33), after subtraction of the flat-space contribution, gives a divergent contribution equal to $(m^2 H^2/16\pi^2)\log(\Lambda_c/m)$ (consistently with the fact that, for $m = 0$, it must vanish because of conformal invariance). After renormalization, this gives a contribution to the vacuum energy which is of order $(m^2 H^2/16\pi^2)\log(M/m)$, and which therefore for $m \ll M$, is negligible with respect to the bosonic contribution (25). It is also interesting to observe that, even in a theory with exact supersymmetry, the cancellation between fermionic and bosonic divergences only takes place at the level of the quartic divergence. There is no contribution proportional to Λ_c^2 from the fermionic sector, and therefore the whole contribution proportional to $H^2\Lambda_c^2$ comes from the bosonic sector. This means that eq. (25) holds even in a theory with exact or broken supersymmetry.

In contrast, gravitons give a contribution to the bare energy density proportional to $\Lambda_c^2 H^2$. In fact, in a FRW space-time each of the two helicity modes $h_{\alpha,k}$, with $\alpha = \{+, \times\}$ satisfies separately the same wave equation as eq. (12) with ϕ_k replaced by $h_{\alpha,k}$,

$$h_{\alpha,k}'' + 2\frac{a'}{a} h_{\alpha,k}' + k^2 h_{\alpha,k} = 0\,.$$
(34)

So each of the two helicity modes gives the same contribution to $\rho_{\rm vac}$ and $p_{\rm vac}$ as a minimally coupled massless scalar field. Therefore, in a theory with n_s minimally coupled elementary scalar fields plus the two degrees of freedom for the graviton, the natural value for the energy density, eq. (25), is of order $(n_s + 2)H^2 M^2/(16\pi^2)$. In particular, even in a theory with no fundamental scalar field, a contribution of order $H^2 M^2/(8\pi^2)$ comes anyhow from gravitons.

B. Vacuum fluctuations during RD and MD

It is straightforward to repeat the same calculation for a radiation-dominated (RD) and for a matter-dominated (MD) era. For the RD epoch we use the modes (16). Then, recalling that $dt = ad\eta$,

$$\dot{\phi}_k = \frac{1}{a}\phi_k' = -\frac{1}{a^2} \left[ik + \frac{a'}{a} \right] e^{-ik\eta}\,,$$
(35)

where as usual the dot denotes the derivative with respect to cosmic time t and the prime the derivative with respect to η. Using $a'/a = \dot{a} = aH(t)$, we get

$$\begin{aligned}
\rho(\Lambda_c) &= \frac{1}{4\pi^2} \int_0^{a\Lambda_c} dk\, k \left(\frac{k^2}{a^4} + \frac{H^2(t)}{2a^2} \right) \\
&= \frac{\Lambda_c^4}{16\pi^2} + \frac{H^2(t)\Lambda_c^2}{16\pi^2}\,,
\end{aligned}$$
(36)

so the energy density turns out to be identical to eq. (22), except that now H is replaced by $H(t)$. For the pressure we get

$$p(\Lambda_c) = \frac{\Lambda_c^4}{48\pi^2} + \frac{H^2(t)\Lambda_c^2}{16\pi^2}\,.$$
(37)

Thus, once removed the Minkowski term, we remain with

$$\rho_{\text{bare}}(\Lambda_c) = \frac{H^2(t)\Lambda_c^2}{16\pi^2}, \qquad (38)$$

while for the pressure we get $p_{\text{bare}}(\Lambda_c) = \rho_{\text{bare}}(\Lambda_c)$. Similarly to what we have discussed in the de Sitter case, this means that the natural value of the renormalized energy density is

$$\rho_{\text{vac}}(t) = \text{const.} \times \sigma \frac{H^2(t)M^2}{16\pi^2}, \qquad (39)$$

where M is the scale where quantum gravity sets in, "const." is a numerical constant $\mathcal{O}(1)$, and $\sigma = \pm 1$. On the other hand, the relation $p_{\text{bare}}(\Lambda_c) = \rho_{\text{bare}}(\Lambda_c)$ does not imply that the renormalized energy density and the renormalized pressure satisfy an equation of state with $w = +1$. As in the de Sitter case, the finite part in the counterterm for the pressure can be chosen so to reproduce any observed value of w, in particular the value $w = -1$. The issue of general covariance is however now more subtle, since even the choice $w = -1$ now leads to

$$\langle 0|T_{\mu\nu}|0\rangle = \text{const.} \times \sigma \frac{H^2(t)M^2}{16\pi^2}(-g_{\mu\nu}) \qquad (40)$$

which, because of the time dependence of $H(t)$, is no longer covariantly conserved. However, general covariance actually only requires that the *total* energy momentum tensor, including that of matter, radiation, etc., be covariantly conserved. The fact that energy-momentum tensor associated to vacuum fluctuations is not separately conserved means that there must be an energy exchange between vacuum fluctuations and other energy sources such as radiation or matter, so that we are actually dealing with an interacting dark energy model. We will come back to this issue in Section IV.

For MD we use the modes given in eq. (15), with $a(\eta) \sim \eta^2$. For the energy density we find

$$\rho(\Lambda_c) = \frac{1}{4\pi^2} \int_0^{a\Lambda_c} dk\, k \left(\frac{k^2}{a^4} + \frac{H^2(t)}{2a^2} + \frac{9H^4(t)}{32k^2} \right). \qquad (41)$$

So, both for MD and for RD, the first two terms are the same as in De Sitter, eq. (22), except that H becomes $H(t)$. Again, the first term is the Minkowski contribution, that we subtract. The term $\sim H^4$ diverges only logarithmically with the UV cutoff (and also require an IR cutoff, which for a light scalar field is provided by its mass, while for a strictly massless field could be taken equal to H). Thus, for the physical, renormalized, energy density we find again that the natural value is given by eq. (39), since the term $\sim H^4$ in eq. (41) produces a subleading term of order $H^4(t)\log(M/H(t))$.

Recalling that (in units $\hbar = c = 1$) Newton's constant is $G = 1/M_{\text{Pl}}^2$, and that the critical density at time t is

$$\rho_c(t) = \frac{3H^2(t)}{8\pi G} = \frac{3}{8\pi}H^2(t)M_{\text{Pl}}^2, \qquad (42)$$

we can express the above result by saying that the "natural" value suggested by QFT is

$$\rho_{\text{vac}}(t) = \text{const.} \times \sigma \frac{1}{6\pi} \left(\frac{M}{M_{\text{Pl}}} \right)^2 \rho_c(t), \qquad (43)$$

with const. $= \mathcal{O}(1)$ and $\sigma = \pm 1$. We therefore find that, both during RD and MD, the renormalized energy density due to zero-point quantum fluctuations is not a constant, and therefore does not contribute to the cosmological constant. Rather, it is a fixed fraction of the critical density $\rho_c(t)$.

As before, each helicity mode of a graviton contributes as a minimally coupled scalar field, while gauge bosons do not contribute and the contribution of light fermions is suppressed by a factor $(m/M)^2 \log(M/m)$. Thus, in a theory with n_s fundamental minimally coupled scalar fields plus the two helicity modes of the graviton we have, for either De Sitter, RD or MD,

$$\rho_Z(t) = \Omega_Z \rho_c(t), \qquad (44)$$

where the subscript Z stands for "zero-point quantum fluctuations", and the natural value of Ω_Z is

$$\Omega_Z \simeq \sigma \frac{(n_s + 2)}{6\pi} \left(\frac{M}{M_{\text{Pl}}} \right)^2, \qquad (45)$$

where M is the scale where quantum gravity sets in, e.g. M_{Pl} itself or the string scale. As we have seen in eqs. (31) and (33), bosons or fermions with mass m not far from the quantum gravity scale M would give corrections of order $[1+\mathcal{O}(m^2/M^2)]$ to this estimate (which could become numerically important if there were, e.g. many fermions at a scale, such as the GUT scale, not far from the Planck mass). The computation of the pressure during MD gives

$$p(\Lambda_c) = \frac{1}{4\pi^2} \int_0^{a\Lambda_c} dk\, k \left(\frac{k^2}{3a^4} + \frac{H^2(t)}{3a^2} + \frac{9H^4(t)}{32k^2} \right). \qquad (46)$$

Again, the first term is the Minkowski contribution, that we subtract, and we neglect the term $\sim H^4$. Thus, we finally find

$$p_{\text{bare}}(\Lambda_c) = \frac{H^2(t)\Lambda_c^2}{24\pi^2}, \qquad (47)$$

and therefore, during MD, $p_{\text{bare}} = +(2/3)\rho_{\text{bare}}$, but again the renormalized energy and pressure satisfy $p = w\rho$ with w determined by the observation.[43]

A technical point that deserves some comment is the choice of the modes given in eqs. (15) and (16). These modes are particularly natural since in the UV limit they reduce to positive-frequency plane waves in flat space. However, the choice of the modes is equivalent to the choice of a particular vacuum state, and the most general possibility is a superposition of positive- and negative-frequency modes with Bogoliubov coefficients α_k and β_k. As we show in appendix B, for a generic choice of vacuum

the dependence of the natural value of $\rho_{\rm vac}$ on $H^2(t)M^2$ is not altered, while the numerical coefficient in front of it can change.

A conceptually interesting aspect of the result (39) is that it appears to involve a mixing of ultraviolet and infrared physics, since it depends both on the UV scale M, and on the horizon size H^{-1}, which represents the "size of the box", and therefore plays the role of an IR cutoff. This is an interesting result by itself, since in quantum field theory we are rather used to the fact that widely separated energy scales decouple. Observes that this UV-IR mixing comes out only because our classical subtraction procedure based on eq. (5) eliminates the troublesome term diverging as Λ_c^4, which is instead a purely UV term. As we already discussed in the Introduction, the origin of this UV-IR mixing can be traced to the fact that, even if $T_{\mu\nu}(x)$ is a local quantity, the vacuum expectation value $\langle 0|T_{\mu\nu}|0\rangle$ is sensitive to $H(t)$ through the definition of the vacuum state $|0\rangle$. In fact, the vacuum is defined from the condition $a_{\mathbf{k}}|0\rangle = 0$, where the annihilation operators $a_{\mathbf{k}}$ are defined with respect to a set of mode functions $\phi_{\mathbf{k}}(t)$; these mode functions are obtained solving a wave equation over the whole space-time, and therefore are sensitive to the time evolution of the scale factor and, more generally, to the overall geometry of space-time.

IV. COSMOLOGICAL IMPLICATIONS

A. Vacuum fluctuations and the dominant component of dark energy

The results of the previous section show that, with the subtraction that we advocate, zero-point vacuum fluctuations do not contribute to the cosmological constant since their energy density, given in eqs. (44) and (45), is not a constant, but rather a fixed fraction of the critical energy at any epoch.

The first question to be addressed is whether a vacuum energy density with such a time behavior could be identified with the dark energy component Ω_Λ which is responsible for the observed acceleration of the universe. If this were the case, since we have found that the vacuum energy density scales as $H^2(t)$, we would have $\Omega_\Lambda(t) = \Omega_\Lambda H^2(t)/H_0^2$ where we follow the standard use of denoting by Ω_Λ the value of $\Omega_\Lambda(t)$ at the present time $t = t_0$. Such a model has been compared to CMB+BAO+SN data in ref. [17], where it is found that it is ruled out at a high significance level.

Formally the result of ref. [17] comes from the fact that the χ^2 obtained fitting this model to the usual estimators for CMB, for BAO and to SNIa turns out to be unacceptably high. However, independently of technical details, it is easy to understand physically why such a model for dark energy is not viable. Consider in fact a spatially flat model with vacuum energy density $\rho_\Lambda(t)$ evolving as $H^2(t)$, with equation of state $w_\Lambda = -1$, and matter density $\rho_M(t)$, with $\Omega_M + \Omega_\Lambda = 1$, in the recent universe

where we can neglect Ω_R. The total energy-momentum conservation is

$$\dot\rho_M + \dot\rho_\Lambda + 3H\rho_M = 0\,, \qquad (48)$$

and cannot be split into two separate conservation equations for ρ_Λ and ρ_M, $\dot\rho_M + 3H\rho_M = 0$ and $\dot\rho_\Lambda = 0$, since the second equation is obviously incompatible with $\rho_\Lambda(t) \sim H^2(t)$, in the recent universe where the Hubble parameter is certainly not a constant. This means that energy must be transferred between $\rho_\Lambda(t)$ and $\rho_M(t)$ in order to obtain the behavior $\rho_\Lambda(t) \sim H^2(t)$, so we have an interacting dark energy model. The Friedmann equation in this model reads

$$
\begin{aligned}
\frac{H^2(t)}{H_0^2} &= \Omega_\Lambda(t) + \Omega_M(t) \\
&= \Omega_\Lambda \frac{H^2(t)}{H_0^2} + \Omega_M(t)\,, \qquad (49)
\end{aligned}
$$

and therefore

$$\frac{H^2(t)}{H_0^2} = \frac{1}{1-\Omega_\Lambda}\Omega_M(t) \qquad (50)$$

and

$$\Omega_\Lambda(t) = \frac{\Omega_\Lambda}{1-\Omega_\Lambda}\Omega_M(t)\,. \qquad (51)$$

In this model, therefore, the time evolution of dark energy density is the same as that of matter. This is clearly incompatible with the existing cosmological observations, that rather indicate that in the recent epoch $\Omega_\Lambda(t)$ remained constant, at least to a first approximation, while the matter density in the concordance ΛCDM model evolves in a way that cannot differ too much from $\Omega_M(t) \sim 1/a^3(t)$. More quantitatively, combining the Friedmann equation (50) with the total energy-momentum conservation equation (48) it can be easily found [17] that in this model the dependence of $\Omega_M(t)$ on the scale factor $a(t)$ is

$$\Omega_M(t) = \frac{\Omega_M}{a^{3-3\Omega_\Lambda}}\,, \qquad (52)$$

where we normalize the scale factor $a(t)$ so that $a(t_0) = 1$ at the present epoch t_0 (we will see explicitly in Sect. IV C how to derive this result in a similar setting). Physically this result expresses the fact that in this model $\rho_\Lambda(t) \sim H^2(t)$ decreases with time, rather than remaining constant as it would do in isolation, and therefore part of its energy density must be transferred to matter, through the energy-conservation equation (48). So, matter energy density decreases more slowly that $1/a^3$. For $\Omega_\Lambda(t)$, recalling that we are considering a flat model with $\Omega_\Lambda + \Omega_M = 1$, eq. (51) gives

$$\Omega_\Lambda(t) = \frac{\Omega_\Lambda}{a^{3-3\Omega_\Lambda}}\,. \qquad (53)$$

To compare these results to observations, we recall that the possibility of a time evolution of the dark energy density is usually studied in the literature by parametrizing it as

$$\Omega_\Lambda(t) = \frac{\Omega_\Lambda}{a^{3+3w_\Lambda(t)}} . \qquad (54)$$

It is important to stress that eq. (54) is simply an effective way of parametrizing the time dependence of $\rho_{\rm DE}$ in order to compare it with the observations, and the parameter w_Λ that appears there is equal to the w parameter that enters in the equation of state of dark energy only if dark energy is not interacting which, as we have seen, is not the case for a model where $\rho_\Lambda(t) \sim H^2(t)$.

Comparing eqs. (53) and (54) we see that a model with $\Omega_\Lambda(t) = \Omega_\Lambda H^2(t)/H_0^2$ predicts that the effective parameter w_Λ in eq. (54) is constant in time, and equal to $-\Omega_\Lambda$. The limit on constant w_Λ obtained from WMAP 7yr data+BAO+SN is $w_\Lambda = -0.980 \pm 0.053$ at 68% c.l. [18] (actually, this value does not include systematic errors in supernova data; including the systematics, the error on w_Λ rather becomes about 0.08, see [19]), and therefore reproducing this value would require $\Omega_\Lambda = +0.980 \pm 0.053$, and $\Omega_M = 1 - \Omega_\Lambda$ consistent with zero to a few percent! It is no wonder that such a model does not fit (any) other cosmological observation. Alternatively, setting $\Omega_M \simeq 0.26$ and $\Omega_\Lambda \simeq 0.74$, we get a prediction $w_\Lambda \simeq -0.74$, which is excluded at a high confidence level.

Therefore, it appears that the interpretation of zero-point quantum fluctuations as the dominant dark energy component, responsible for the acceleration of the universe, is not viable. In the next section we will explore the possibility that they could still provide a new, subdominant, "dark" component, that we will denote by ρ_Z to distinguish it from the dominant component ρ_Λ which instead, according to the limits on w_Λ, is at least approximately constant in time. We will see that, for plausible values of the mass scale M where quantum gravity sets in, the energy density ρ_Z can have a value consistent with existing observations, but still potentially detectable.

B. Theoretical expectations for Ω_Z

The effect of ρ_Z on the cosmological expansion depends on the mass scale M, whose exact value can only be determined once one has a fundamental theory of quantum gravity. If we set $M = M_{\rm Pl}$ and we consider the Standard Model with one Higgs field, so $n_s = 1$, eq. (45) gives $|\Omega_Z| \simeq 1/(2\pi) \simeq 0.16$. However, precise numerical factors are beyond such order-of-magnitude estimates and, by lowering the UV scale M, it is easy to reduce this number to smaller but still potentially observable values. For instance, in heterotic string theory the scale is rather given by the heterotic string mass scale $M_H = gM_{\rm Pl}$, where $g \simeq 1/5$ is the value of the gauge couplings at the string scale [20]. This would rather lead to the estimate $|\Omega_Z| \simeq (n_s + 2)g^2/(6\pi) \simeq 2 \times 10^{-3}(n_s + 2)$. Lower values of the cutoff, possibly down to the TeV scale, can

be obtained in theories with large extra dimensions [21]. It is also important to observe that, as explained in the discussion below eq. (33), the estimate given in eq. (45) holds even in a theory with exact or broken supersymmetry, since the contribution proportional to H^2M^2 comes anyhow only from the bosonic sector.

C. Cosmological evolution equations

We consider a flat ΛCDM cosmology, with a vacuum energy Ω_Λ having an equation of state $p_\Lambda = w_\Lambda\rho_\Lambda$ and we further add the energy density $\rho_Z(t)$ given in eq. (44), with $p_Z = w_Z\rho_Z$. We have nothing to add to the problem of the physical origin of ρ_Λ, except that in our model it is not due to zero-point quantum fluctuations: it has nothing to do with the quartic divergence in the vacuum energy (which is eliminated by our ADM-like subtraction), nor with the quadratic divergence, which is instead the origin of ρ_Z. Then in this model (which could be conveniently called ΛZCDM)

$$\rho = \rho_R + \rho_M + \rho_\Lambda + \rho_Z , \qquad (55)$$

$$p = \frac{1}{3}\rho_R + w_\Lambda\rho_\Lambda + w_Z\rho_Z . \qquad (56)$$

The values $w_Z = w_\Lambda = -1$ will be assumed in the following (in appendix C we discuss the case w_Z generic). From eq. (44),

$$\rho_Z(t) = \Omega_Z\rho_c(t) = \left(\frac{\Omega_Z\rho_0}{H_0^2}\right) H^2(t) , \qquad (57)$$

where $\rho_0 = 3H_0^2/(8\pi G)$ is the present value of the critical density, so the energy density in the dark sector is

$$\rho_{\rm DE} \equiv \rho_\Lambda + \rho_Z$$
$$= \rho_\Lambda + \left(\frac{3\Omega_Z}{8\pi G}\right) H^2(t) . \qquad (58)$$

Quite interestingly, this is the same form of the dark energy density found in refs. [22–25] from an apparently rather different approach, namely from the suggestion that the cosmological constant could evolve under renormalization group, after identifying their parameter ν with our Ω_Z, compare with eq. (13) of ref. [23]. Observe also that their value for the parameter ν is $\nu = \pm(1/12\pi)M^2/M_{\rm Pl}^2$ where M is the mass scale where new physics comes in. This has the same parametric dependence on $(M/M_{\rm Pl})$ as our result for Ω_Z, and is even quite close numerically [44].

Using eq. (58), the Friedmann equation becomes

$$H^2(t) = \frac{8\pi G}{3}\left[\rho_R + \rho_M + \rho_\Lambda + \left(\frac{3\Omega_Z}{8\pi G}\right)H^2(t)\right] , \qquad (59)$$

which can be rewritten as

$$H^2(t) = \frac{H_0^2}{1-\Omega_Z}[\Omega_R(t) + \Omega_M(t) + \Omega_\Lambda(t)] , \qquad (60)$$

where $\Omega_i(t) = \rho_i(t)/\rho_0$, $(i = R, M, \Lambda)$. Thus, the effect of Ω_Z on the Friedmann equation is equivalent to a rescaling of the present value of the Hubble constant, $H_0 \rightarrow H_0/(1 - \Omega_Z)^{1/2}$ or, equivalently, to a rescaling of $\Omega_i(t)$ (with $i = R, M, \Lambda$) into $\Omega_i(t)/(1 - \Omega_Z)$. Even if we set $w_\Lambda = -1$ we have for the moment written $\Omega_\Lambda(t)$, rather than setting it to a constant, in order to allow for the possibility of an energy exchange between $\rho_Z(t)$ and $\rho_\Lambda(t)$, see below.

Zero-point fluctuations also contribute to the equation for the acceleration

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p), \tag{61}$$

in a way which depends on the sign of ρ_Z. With $w_Z = -1$ (or more generally whenever $w_Z < -1/3$), zero-point fluctuations contribute to accelerating the universe if $\rho_Z > 0$, while for $\rho_Z < 0$ they give a contribution that decelerates the expansion, and which therefore opposes the accelerating effect of ρ_Λ.

We next consider the energy conservation equation. As we already mentioned in Section III B, the fact that zero-point quantum fluctuations have an energy density $\rho_Z(t) \sim H^2(t)$ has non-trivial consequences on the conservation of energy. Consider in fact the energy conservation in a FRW background,

$$\dot{\rho} = -3H(\rho + p). \tag{62}$$

When there is no exchange of energy among the different components, energy conservation is satisfied separately for each component, so $\dot{\rho}_i = -3H(\rho_i + p_i)$ with $i = R, M, \Lambda, Z$. If this were the case, using $p_Z = w_Z \rho_Z$, the equation

$$\dot{\rho}_Z = -3H(\rho_Z + p_Z) = -3(1 + w_Z)H\rho_Z \tag{63}$$

would give $\rho_Z(t) \sim a^{-3(1+w_Z)}$, and in particular $\rho_Z(t)$ constant if $w_Z = -1$. However, we have found that $\rho_Z(t) \sim H^2(t)$. So, energy must be transferred between zero-point fluctuations and other components [45].

In principle, there are various mechanisms by which vacuum fluctuations can exchange energy with ordinary matter. A typical example is the amplification of vacuum fluctuations [27, 28], or the change in a large-scale scalar field due to the continuous flow across the horizon of small-scale quantum fluctuations of the same scalar field, which is also at the basis of stochastic inflation [29]. If we assume that ρ_Z exchanges energy with ρ_M but not with ρ_Λ the relevant conservation equation is (setting hereafter $w_Z = -1$)

$$\dot{\rho}_M + \dot{\rho}_Z = -3H\rho_M. \tag{64}$$

Using eq. (57), we then obtain

$$\dot{\rho}_M = -3H\rho_M - \left(\frac{\Omega_Z \rho_0}{H_0^2}\right)\frac{dH^2(t)}{dt}. \tag{65}$$

This equation was already discussed in refs. [17, 24] in the context of their time-varying cosmological constant

model and, following these papers, to solve it we compute the term $dH^2(t)/dt$ on the right-hand side by using the Friedmann equation (60) and we obtain (neglecting for simplicity $\Omega_R(t)$ in the low redshift epoch in which we are interested here)

$$\dot{\rho}_M = -3H\rho_M - \frac{\Omega_Z}{1 - \Omega_Z}\dot{\rho}_M, \tag{66}$$

where we used the fact that, since we are assuming that ρ_Z only interacts with ρ_M, and we are furthermore assuming $w_\Lambda = -1$, the energy density ρ_Λ evolves in isolation and satisfies $\dot{\rho}_\Lambda = 0$. Equation (66) can be rewritten as

$$\dot{\rho}_M = -3(1 - \Omega_Z)H\rho_M, \tag{67}$$

and has the solution

$$\rho_M(z) = \rho_M(0)(1 + z)^{3(1-\Omega_Z)}. \tag{68}$$

Of course, since we have taken $w_Z = -1$, if vacuum energy density were non-interacting it would remain constant in time. In the case $\sigma = +1$ (i.e. when $\Omega_Z > 0$) we have rather found that ρ_Z is proportional to $+H^2(t)$, and therefore it decreases with time instead of staying constant. This means that dark energy is interacting, and that energy is transferred from vacuum fluctuations to matter, for instance with a mechanism analogous to the amplification of vacuum fluctuations, and as a result the energy density of matter must decrease slower than $1/a^3$. This is reflected in eq. (68), since for $\Omega_Z > 0$ we find that ρ_M indeed decreases slower than $1/a^3$. If $\Omega_Z < 0$ the situation is reversed. A behavior $\rho_Z(t) \sim -H^2(t)$ means that $\rho_Z(t)$ becomes less and less negative as time increases, so energy is transferred from matter to vacuum fluctuations, and $\Omega_M(z)$ decreases faster than $1/a^3$.

In the absence of an understanding of the dynamical origin of the dominant dark energy term ρ_Λ, it is interesting to consider also the possibility that energy could be exchanged also between ρ_Z and ρ_Λ. We assume at first, for simplicity, that energy is exchanged only with ρ_Λ, and not with ρ_M. Then the corresponding conservation equation (taking $w_\Lambda = -1$ for simplicity) is

$$\dot{\rho}_\Lambda + \dot{\rho}_Z = 0, \tag{69}$$

which trivially integrates to

$$\rho_\Lambda(t) + \rho_Z(t) = \text{constant} = \rho_\Lambda(t_0) + \rho_Z(t_0) \tag{70}$$

and therefore, using $\rho_Z(t) = \Omega_Z \rho_0 H^2(t)/H_0^2$,

$$\Omega_\Lambda(z) = \Omega_\Lambda - \Omega_Z\left[\frac{H^2(z)}{H_0^2} - 1\right], \tag{71}$$

where, as usual, on the right-hand side $\Omega_\Lambda \equiv \Omega_\Lambda(z = 0)$. However, in terms of the total dark energy density defined in eq. (58), we see that in this case we simply have a total dark energy density ρ_{DE} that satisfies $\dot{\rho}_{\text{DE}} = 0$,

while the matter energy density satisfies its usual conservation equation $\dot{\rho}_M + 3H\rho_M = 0$, and also eqs. (59) and (61), when rewritten in terms of ρ_{DE}, take the standard ΛCDM form. Thus, in the end a model where ρ_Z only exchanges energy with ρ_Λ (and in which $w_Z = w_\Lambda = -1$) is indistinguishable from standard ΛCDM cosmology with just a cosmological constant. In the most general case in which $w_Z \neq -1$, however, even setting $w_\Lambda = -1$ the pressure $p_{DE} = p_Z + p_\Lambda = w_Z\rho_Z + w_\Lambda p_\Lambda$ is no longer equal to $-\rho_{DE}$, so this model has observable deviations for ΛCDM, and is in fact of the type called ΛXCDM [30–33]. In the following we will however restrict to $w_Z = w_\Lambda = -1$.

A more general phenomenological analysis, in which one takes into account the possibility that ρ_Z interacts both with ρ_Λ and with ρ_M, can be performed by splitting the conservation equation

$$\dot{\rho}_M + \dot{\rho}_\Lambda + \dot{\rho}_Z = -3H\rho_M \qquad (72)$$

into the two equations

$$\dot{\rho}_\Lambda = -(1-\alpha)\dot{\rho}_Z , \qquad (73)$$
$$\dot{\rho}_M = -3H\rho_M - \alpha\dot{\rho}_Z , \qquad (74)$$

where $0 \leq \alpha \leq 1$. Equation (64) corresponds to the limiting case $\alpha = 1$ while eq. (69) corresponds to the limiting case $\alpha = 0$. Rewriting these equations in terms of the total dark energy density $\rho_{DE} = \rho_\Lambda + \rho_Z$, we get

$$\dot{\rho}_{DE} = \alpha\dot{\rho}_Z , \qquad (75)$$
$$\dot{\rho}_M = -3H\rho_M - \alpha\dot{\rho}_Z , \qquad (76)$$

which shows that the observable consequences depend only on the combination $\alpha\rho_Z$. In other words, as long as $w_Z = w_\Lambda = -1$, there is no point in postulating an energy exchange between ρ_Z and ρ_Λ, since only the fraction of ρ_Z which is exchanged with ρ_M has observable consequences. In the following we will limit ourselves to $w_Z = w_\Lambda = -1$, and we then set $\alpha = 1$.

D. Limits on Ω_Z from cosmological observations

In this section we perform a first analysis of the limits that some cosmological observations impose on Ω_Z. A more detailed comparison with the data will be presented elsewhere. The fact that the energy budget of the universe at the present epoch is known to a precision of about 1% by itself does not yet constraint Ω_Z, since a part of what is normally attributed to Ω_Λ could be due to Ω_Z. The only way of disentangling them is by using their different temporal evolution, since $\rho_Z(t) \sim H^2(t)$ while the dominant component ρ_Λ is constant, at least within the present experimental accuracy. In the next subsections we examine various limits on Ω_Z which make use of the time dependence $\rho_Z(t) \sim H^2(t)$.

1. Bound on Ω_Z from BBN

We first examine the bound coming from big-bang nucleosynthesis (BBN), which constraints the energy budget of the universe at that epoch. The limit on extra contributions to the energy density at time of BBN is usually expressed in terms of the effective number of neutrino species N_ν, defined so that any extra source of energy density, compared to the Standard Model, is written as

$$\frac{\rho_{extra}}{\rho_\gamma} = \frac{7}{8}\Delta N_\nu , \qquad (77)$$

and $\Delta N_\nu = N_\nu - N_\nu^{SM}$, where $N_\nu^{SM} \simeq 3.046$ is the value predicted by the Standard Model with three light neutrino families, after taking into account finite temperature QED corrections and the fact that neutrino decoupling is not instantaneous [34]. The most recent BBN bound is $N_\nu \leq 3.6$ at 95% c.l. [35], corresponding to a limit $\Delta N_\nu \leq 3.6 - 3.046 \simeq 0.55$. At the epoch of BBN only the photons and the three neutrinos contribute significantly to the energy density, while $e^\pm$ already annihilated into photons, resulting in a photon temperature higher than the neutrino temperature by a factor $(11/4)^{1/3}$. Therefore, at BBN,

$$\rho_c = \rho_\gamma \left[1 + 3 \times \frac{7}{8}\left(\frac{4}{11}\right)^{4/3}\right] , \qquad (78)$$

where the factor 7/8 comes from Fermi statistics. Combining this with eq. (77) gives

$$\left(\frac{\rho_{extra}}{\rho_c}\right)_{BBN} = \frac{(7/8)\Delta N_\nu}{1 + 3 \times \frac{7}{8}\left(\frac{4}{11}\right)^{4/3}} \leq 0.29 , \qquad (79)$$

This gives a corresponding bound on the total dark energy density $\rho_{DE} = \rho_\Lambda + \rho_Z$ at the time of nucleosynthesis. From eq. (75) with $\alpha = 1$ we have $\dot{\rho}_{DE} = \dot{\rho}_Z$, so

$$\rho_{DE}(z) = \rho_Z(z) + [\rho_{DE}(0) - \rho_Z(0)] . \qquad (80)$$

At $z = z_{BBN}$ the constant term in bracket is totally negligible with respect to the critical density $\rho_c(z_{BBN})$, while $(\rho_Z/\rho_c)_{BBN} = \Omega_Z$, so the bound (79) translates into

$$\Omega_Z < 0.29 . \qquad (81)$$

Strictly speaking, BBN gives an upper bound only on Ω_Z, and not on its absolute value, since negative values of ρ_Z could in principle be compensated by other forms of energy.

2. Bound on Ω_Z from CMB+BAO+SNIa

A comparison of this model to CMB+BAO+SNIa data has been performed in ref. [17] within the context of the model of a cosmological constant that evolves under renormalization group [22–25]. Even if our physical

motivations are different from theirs, the model is phenomenologically the same, and our parameter Ω_Z corresponds to the parameter that they denote as γ (or as ν). We can therefore translate immediately their result in our setting. In particular, from a joint analysis of CBM+BAO+SNIa, sampling the interval $\Omega_Z \in [0, 0.3]$ [46] in steps of 0.001, they find a best fit value $\Omega_Z = 0.002 \pm 0.001$.

3. Bound on Ω_Z from the limits on the time evolution of dark energy

It is also instructive to compare the time evolution of a dark energy density of the form $\rho_{\mathrm{DE}}(z) = \rho_\Lambda + \rho_Z(z)$ to the existing observational limit, which is also obtained from a combination of CMB+BAO+SNIa data.

In standard ΛCDM cosmology, without the contribution ρ_Z, the dark energy density is given uniquely by ρ_Λ and, allowing for a generic w_Λ, its evolution with redshift z is given by

$$\rho_\Lambda(z) = \rho_\Lambda(0)(1+z)^{3+3w_\Lambda}, \qquad (82)$$

while (neglecting Ω_R, since we are interested here in the evolution at small redshifts) the critical density $\rho_c(z)$ is given by

$$\rho_c(z) = \rho_c(0)[\Omega_M(1+z)^3 + \Omega_\Lambda], \qquad (83)$$

with $\Omega_\Lambda \simeq 0.738$. Then

$$\frac{\rho_\Lambda(z)}{\rho_c(z)} = \frac{\Omega_\Lambda(1+z)^{3+3w_\Lambda}}{\Omega_M(1+z)^3 + \Omega_\Lambda}. \qquad (84)$$

We use the recent determination of w_Λ given in [19], $w_\Lambda = -0.997^{+0.077}_{-0.082}$ at 68% c.l., which includes also systematic errors on supernova data (the more stringent bound $w_\Lambda = -0.980 \pm 0.053$ given in [18] only includes the statistical error in the SN data). In Fig. 1 we then plot the function $\rho_\Lambda(z)/\rho_c(z)$ given in eq. (84), in correspondence of the 1σ upper and lower limits on w_Λ, $(w_\Lambda)_{\max} = -0.997 + 0.077$ and $(w_\Lambda)_{\min} = -0.997 - 0.082$, respectively (black solid lines). We limit ourselves to the redshift interval $0 \le z \le 0.5$ that, as shown in [19, 36], is responsible for most part of the bound on w_Λ. Plotting the constraint on w_Λ, or the corresponding constraints on $\rho_\Lambda(z)$, in different redshift bins, one finds in fact that the bin $0.5 \le z \le 1$ already gives a poorly constrained w, see e.g. Fig. 15 of [19].

In this range of redshifts, we compare the temporal evolution given in eq. (84) with $w_\Lambda = (w_\Lambda)_{\min}$ and with $w_\Lambda = (w_\Lambda)_{\max}$, respectively, to that obtained in ΛZCDM (using for definiteness the values $w_\Lambda = w_Z = -1$). From eq. (80),

$$\rho_{\mathrm{DE}}(z) = \Omega_Z \rho_c(z) + \rho_0(\Omega_{\mathrm{DE}} - \Omega_Z), \qquad (85)$$

where $\Omega_{\mathrm{DE}} \equiv \Omega_\Lambda + \Omega_Z \simeq 0.738$. To write explicitly the critical density $\rho_c(z) = \rho_0 H^2(z)/H_0^2$ we use the Friedmann equation (60) (neglecting again Ω_R) together with
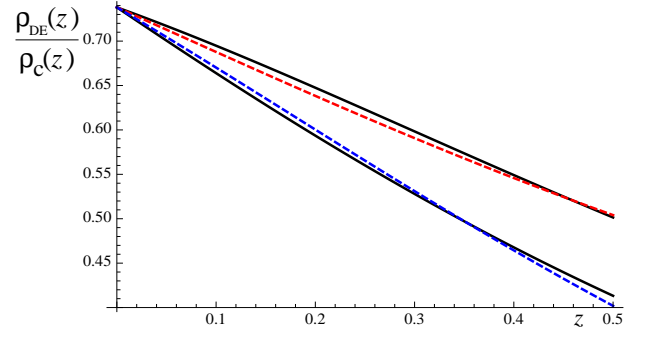

FIG. 1: The functions $\rho_\Lambda(z)/\rho_c(z)$ for $w_\Lambda = (w_\Lambda)_{\min}$ and $(w_\Lambda)_{\max}$ (black solid lines) compared to the function $\rho_{\mathrm{DE}}(z)/\rho_c(z)$ given by eq. (88) with $\Omega_Z = +0.10$ (upper dashed curve, red) and $\Omega_Z = -0.10$ (lower dashed curve, blue).

eq. (68), so

$$\begin{aligned}\frac{H^2(z)}{H_0^2} &= \frac{1}{1-\Omega_Z}[\Omega_M(z) + \Omega_\Lambda] \\ &= \frac{1}{1-\Omega_Z}[\Omega_M(1+z)^{3(1-\Omega_Z)} + \Omega_\Lambda], \quad (86)\end{aligned}$$

and therefore, writing $\Omega_\Lambda = \Omega_{\mathrm{DE}} - \Omega_Z$,

$$\rho_c(z) = \rho_0 \frac{\Omega_M(1+z)^{3(1-\Omega_Z)} + \Omega_{\mathrm{DE}} - \Omega_Z}{1 - \Omega_Z}. \qquad (87)$$

Combining eqs. (85) and (87) we get

$$\frac{\rho_{\mathrm{DE}}(z)}{\rho_c(z)} = \Omega_Z + \frac{(\Omega_{\mathrm{DE}} - \Omega_Z)(1 - \Omega_Z)}{\Omega_M(1+z)^{3(1-\Omega_Z)} + \Omega_{\mathrm{DE}} - \Omega_Z}. \qquad (88)$$

This function is plotted in Fig. 1, keeping Ω_{DE} fixed at the observed value $\Omega_{\mathrm{DE}} \simeq 0.738$, and choosing $\Omega_Z = +0.10$ (upper dashed line, red) and $\Omega_Z = -0.10$ (lower dashed line, blue). We see that, at least at this relatively crude level of analysis, values of Ω_Z in the approximate range $|\Omega_Z| \lesssim 0.1$ are consistent with the observational limits on the temporal evolution of dark energy, since the corresponding curves stay inside the two curves with $w_\Lambda = (w_\Lambda)_{\min}$ and $w_\Lambda = (w_\Lambda)_{\max}$, respectively, down to the maximum redshifts $z \simeq 0.5$ where $\rho_\Lambda(z)$ is significantly constrained by the data. If one would rather compare the function $\rho_\Lambda(z)/\rho_c(z)$ given in eq. (84), to the 3σ upper and lower limits on $(w_\Lambda)_{\max} = -0.997 + 3 \times 0.077$ and $(w_\Lambda)_{\min} = -0.997 - 3 \times 0.082$, one would rather find $-0.20 \lesssim \Omega_Z \lesssim 0.35$. Of course this analysis gives only a first rough but intuitive estimate of the bound that can be obtained from the limits on the redshift dependence of ρ_{DE}. A more accurate study requires fitting this model to the data, as in [17].

E. How not to solve the cosmological constant problem

We think that another useful aspect of the above analysis is to put in a sharper focus where the main difficulty is, in explaining the observed value of dark energy density. If one looks at eq. (39), which holds both in RD and in MD, setting t equal to the present time t_0 and $M \simeq M_{\rm Pl}$, one finds that the energy density associated to vacuum fluctuations today is $\rho_Z(t_0) \sim H_0^2 M_{\rm Pl}^2$, which is of the right order of magnitude of the observed dark energy density (it could even be tempting to observe, from eqs. (44) and (45), that with $M = M_{\rm Pl}$, $n_s = 12$ and $\sigma = +1$ one gets $\Omega_Z \simeq 7/(3\pi) \simeq 0.743$, which is very close to the measured value $\Omega_\Lambda \simeq 0.738$). However, at this stage this observation is not yet a possible explanation of the numerical value of the cosmological constant, not even at the level of orders of magnitude. The trouble is that the same computation, performed at a generic time $t \neq t_0$, gives $\rho_Z(t) \sim H^2(t)M_{\rm Pl}^2$, so the resulting energy density is not constant. As we have discussed in Section IV A, such a time behavior is observationally excluded, at least for the dominant dark energy component, and can only be accepted for a suitably small subdominant dark component.

We should observe that the same conclusion also applies to some existing attempts at computing the cosmological constant which make use of H_0 and $M_{\rm Pl}$, such as the holographic approach to the cosmological constant [37–39], where again one obtains a value of order $M_{\rm Pl}^2 H_0^2$ today. However, the very same reasoning would give $M_{\rm Pl}^2 H^2(t)$ at a generic time, which as we have seen is ruled out, at least for the dominant component of dark energy.

A similar remark can also be made for the result of ref. [40], where it is proposed that the trace anomaly in QCD gives a contribution to the vacuum energy density proportional to $\Lambda_{\rm QCD}^3$ times the Hubble parameter to the first power. Using the present value of the Hubble parameter, ref. [40] finds that $(\Lambda_{\rm QCD}^3 H_0)/\rho_0$ is roughly comparable to Ω_Λ (actually, this is true only within about one or two orders of magnitude; for the typical values of $\Lambda_{\rm QCD} \simeq 100-200\,{\rm MeV}$, we get $(\Lambda_{\rm QCD}^3 H_0)/\rho_0 \simeq 25-200$, not that close to $\Omega_\Lambda = 0.7$. Of course precise numerical factor were anyhow beyond the estimate in ref. [40]). In any case, the suggestion of ref. [40] that this effect has a potential relevance for explaining the observed acceleration of the universe faces a problem similar to the one discussed above. In fact, if at the present time t_0 one finds $\rho_{\rm QCD} \sim \Lambda_{\rm QCD}^3 H_0$, the same calculation, performed at a generic time t of course gives $\rho_{\rm QCD}(t) \sim \Lambda_{\rm QCD}^3 H(t)$. As shown in ref. [17], this behavior is ruled out by the comparison with CMB+BAO+SNIa data. (It should also be observed that a contribution to the vacuum energy density proportional to an odd power of $H(t)$ is not consistent with the general covariance of the effective action for gravity, see Section 3.1 of ref. [14]).

What we learn from the above examples is that the real challenge, in explaining the cosmological constant, is not so much to explain its numerical value today; having at our disposal the two scales $M_{\rm Pl}$ and H_0, once the term proportional to $M_{\rm Pl}^4$ is eliminated one naturally remains with a result proportional to $M_{\rm Pl}^2 H_0^2$, which gives the right order of magnitude. The real challenge is to find a dynamical mechanism that gives a value of order $M_{\rm Pl}^2 H_0^2$ today, without giving $M_{\rm Pl}^2 H^2(t)$ at a generic time t, which is the essence of the coincidence problem.

V. CONCLUSIONS

One aspect of the cosmological constant problem, or more generally of the problem of understanding the origin of dark energy, is to understand why zero-point fluctuations of quantum fields do not produce an energy density of the order of M^4, where M is the UV mass scale of the quantum field theory (e.g. the Planck mass, or the string mass scale), despite the fact that this seems to be the natural value suggested by quantum field theory. We have proposed that the solutions to this long-standing puzzle has a purely classical origin, and is related to the correct definition of energy in classical General Relativity, which already involves the subtraction of the flat-space contribution, see eq. (5).

We have applied this subtraction procedure to a FRW space-time with Hubble parameter $H(t)$ and we have found that the remaining energy density, after renormalization, has a "natural" value proportional to $M^2 H^2(t)$ (and a sign that could in principle be either positive or negative, just as in the Casimir effect). For $M \simeq M_{\rm Pl}$ this gives an energy density just of the order of the critical density of the universe. As we have discussed, however, such an energy density has a time dependence that is not compatible with present observations, if we identify it with the dark energy component with $\Omega_\Lambda \simeq 0.7$ which in the standard ΛCDM cosmology is responsible for the observed acceleration of the universe. It is however possible that it represents a new form of dark energy, whose normalized energy density today, Ω_Z, is smaller than Ω_Λ. Values of $|\Omega_Z| \lesssim$ a few $\times 10^{-3}$ are compatible with the observations that we have discussed, but could give observable effects in more detailed studies that make use of the specific signature of zero-point fluctuations, namely an energy density with a time dependence proportional to $H^2(t)$, as well as in future more accurate cosmological observations.

Appendix A: Computation of ρ and p in FRW background

Even if the computation leading to eqs. (20) and (21) is elementary, we find it useful to report it here. The energy E associated to a real massless scalar field ϕ in a FRW metric, in a comoving volume V, is

$$
\begin{aligned}
E &= \int_V d^3x \sqrt{-g}\, T_{00} \\
&= \int_V d^3x \sqrt{-g} \left[\frac{1}{2}(\partial_0\phi)^2 + \frac{1}{2a^2}(\partial_i\phi)^2 \right] , \quad \text{(A1)}
\end{aligned}
$$

where we used $T_{\mu\nu}$ from eq. (17). Observe that x^i are comoving coordinates and the factor $\sqrt{-g} = a^3$ transforms the comoving volume element d^3x into the physical volume element. Using the mode expansion (11) we get, taking for illustration the term $(\partial_0\phi)^2$,

$$
\begin{aligned}
\int d^3x \sqrt{-g}\,(\partial_0\phi)^2 &= \int \frac{d^3k}{(2\pi)^3\sqrt{2k}} \int \frac{d^3k'}{(2\pi)^3\sqrt{2k'}} \\
&\times \int d^3x \sqrt{-g} \left[a_{\mathbf{k}} \dot{\phi}_k e^{i\mathbf{k}\cdot\mathbf{x}} + a_{\mathbf{k}}^\dagger \dot{\phi}_k^* e^{-i\mathbf{k}\cdot\mathbf{x}} \right] \\
&\times \left[a_{\mathbf{k}'} \dot{\phi}_{k'} e^{i\mathbf{k}'\cdot\mathbf{x}} + a_{\mathbf{k}'}^\dagger \dot{\phi}_{k'}^* e^{-i\mathbf{k}'\cdot\mathbf{x}} \right] .
\end{aligned}
$$

Performing the integral in d^3x over a volume V large compared to the wavelength of all modes of interest we have

$$
\int_V d^3x\, e^{i(\mathbf{k}\pm\mathbf{k}')\cdot\mathbf{x}} = (2\pi)^3 \delta^{(3)}(\mathbf{k}\pm\mathbf{k}') , \quad \text{(A2)}
$$

and we get

$$
\begin{aligned}
E &= \frac{a^3(t)}{2} \int \frac{d^3k}{(2\pi)^3 2k} \left(|\dot{\phi}_k|^2 + \frac{k^2}{a^2}|\phi_k|^2 \right) \left[a_{\mathbf{k}}^\dagger a_{\mathbf{k}} + a_{\mathbf{k}} a_{\mathbf{k}}^\dagger \right] \\
&+ \frac{a^3(t)}{2} \int \frac{d^3k}{(2\pi)^3 2k} \left[\left(\dot{\phi}_k^2 + \frac{k^2}{a^2}\phi_k^2 \right) a_{\mathbf{k}} a_{-\mathbf{k}} + \text{h.c.} \right] . \text{(A3)}
\end{aligned}
$$

Observe that in flat Minkowski space $a(t) = 1$ and (since we are considering a massless field) $\phi_k(t) \sim e^{-iE_k t} = e^{-ikt}$, so the term proportional to $a_{\mathbf{k}} a_{-\mathbf{k}}$ vanishes. In a generic curved space it is instead non-zero, so a general state in a curved background is characterized by the expectation values $\langle a_{\mathbf{k}}^\dagger a_{\mathbf{k}} \rangle$ and $\langle a_{\mathbf{k}} a_{-\mathbf{k}} \rangle$ [8]. For the vacuum state, however, the only non-vanishing contribution comes from the term proportional to $a_{\mathbf{k}} a_{\mathbf{k}}^\dagger$ in eq. (A3), and can be computed using $[a_{\mathbf{k}}, a_{\mathbf{k}'}^\dagger] = (2\pi)^3 \delta^{(3)}(\mathbf{k}-\mathbf{k}')$, from which it follows that $[a_{\mathbf{k}}, a_{\mathbf{k}}^\dagger] = V$, where V is the comoving spatial volume (since $\mathbf{k}$ is a comoving momentum), and therefore

$$
E_{\text{vac}} = \frac{1}{2} V a^3(t) \int \frac{d^3k}{(2\pi)^3 2k} \left(|\dot{\phi}_k|^2 + \frac{k^2}{a^2}|\phi_k|^2 \right) . \quad \text{(A4)}
$$

Multiplying the comoving volume V by the factor $\sqrt{-g} = a^3(t)$ we recover the physical volume V_{phys}, and therefore

the energy of zero-point quantum fluctuations is

$$
E_{\text{vac}} = \frac{V_{\text{phys}}}{2} \int \frac{d^3k}{(2\pi)^3 2k} \left(|\dot{\phi}_k|^2 + \frac{k^2}{a^2}|\phi_k|^2 \right) . \quad \text{(A5)}
$$

The vacuum energy density is then defined as $E_{\text{vac}}/V_{\text{phys}}$.

For the pressure, the spatial isotropy of the FRW metric implies that $p = T_1^1 = T_2^2 = T_3^3$ (observe that, with our signature $(-,+,+,+)$ the energy-momentum tensor of a perfect fluid is $T_\nu^\mu = \text{diag}(-\rho, p, p, p)$). It is convenient to write $p = (1/3)\sum_i T_i^i$ and, as we have done for the energy density, consider first the integrated quantity

$$
\begin{aligned}
P &= \frac{1}{3} \int_V d^3x \sqrt{-g} \sum_i T_i^i \quad \text{(A6)} \\
&= \int_V d^3x \sqrt{-g} \left[\frac{1}{2}(\partial_0\phi)^2 - \frac{1}{6a^2}(\partial_i\phi)^2 \right] ,
\end{aligned}
$$

where, in the second line, the sum over i is understood. Repeating the same steps as above, we get the zero-point contribution

$$
P_{\text{vac}} = \frac{V_{\text{phys}}}{2} \int \frac{d^3k}{(2\pi)^3 2k} \left(|\dot{\phi}_k|^2 - \frac{k^2}{3a^2}|\phi_k|^2 \right) , \quad \text{(A7)}
$$

and $p_{\text{vac}} = P_{\text{vac}}/V_{\text{phys}}$. The off-diagonal elements of the volume integral of $T_{\mu\nu}$ vanish trivially, since they involves integrations over $k_0 k_i$, or over $k_i k_j$ with $i \neq j$, which vanish by parity.

Appendix B: Dependence on the choice of vacuum

A point that deserves some comment is the choice of the modes given in eqs. (15) and (16). These modes are particularly natural since in the UV limit they reduce to positive-frequency plane waves in flat space. However, the choice of the modes is equivalent to the choice of a particular vacuum state, and the most general possibility is a superposition of positive- and negative-frequency modes (15), (16) with Bogoliubov coefficients α_k and β_k,

$$
\phi_k(\eta) = \frac{\alpha_k}{a(\eta)} \left(1 - \frac{i\epsilon}{k\eta} \right) e^{-ik\eta} + \frac{\beta_k}{a(\eta)} \left(1 + \frac{i\epsilon}{k\eta} \right) e^{+ik\eta} ,
\tag{B1}
$$

where $\epsilon = 0$ for RD and $\epsilon = 1$ for De Sitter and MD, and the Bogoliubov coefficient satisfy the normalization condition $|\alpha_k|^2 - |\beta_k|^2 = 1$. It is straightforward to repeat the computation of the vacuum energy density using the modes (B1). When computing $|\phi_k|^2$ and $|\dot{\phi}_k|^2$, mixed term proportional to $\alpha_k \beta_k^*$ and to $\alpha_k^* \beta_k$ have a time dependence which contains the factors $\exp\{\pm 2ik\eta\}$. After integrating over k these produce terms proportional to $\sin(2a\Lambda_c\eta)$ and $\cos(2a\Lambda_c\eta)$. Since

$$
a\Lambda_c\eta = \Lambda_c a(t) \int^t \frac{dt'}{a(t')} = \mathcal{O}(\Lambda_c t) , \quad \text{(B2)}
$$

these terms oscillate very fast in time, with a Planckian frequency, and therefore they average to zero over any

macroscopic time interval, and can be dropped. Keeping only the contributions proportional to $|\beta_k|^2$ and to $|\alpha_k|^2 = 1 + |\beta_k|^2$, subtracting as usual the Minkowski term (and neglecting again the term $\mathcal{O}(H^4)$ which appears in the MD case), for a real scalar field we find

$$\rho_{\text{bare}}(\Lambda_c) = \frac{H^2}{8\pi^2 a^2} \int_0^{a\Lambda_c} dk\, k(2n_k + 1), \qquad (B3)$$

and $p_{\text{bare}}(\Lambda_c) = w_{\text{bare}}\rho_{\text{bare}}(\Lambda_c)$, where w_{bare} is the same found before. Therefore a different choice for the vacuum affects the numerical value of Ω_Z in eq. (44) by a numerical factor which reflects the occupation number of the various modes.

Appendix C: Cosmological equations for w_Z generic

In Section IV C we studied the cosmological evolution equations setting $w_Z = -1$. In this appendix we study how vacuum fluctuations affect the matter evolution for w_Z generic. Then eq. (64) generalizes to

$$\dot{\rho}_M + \dot{\rho}_Z = -3H\rho_M - 3(1 + w_Z)H\rho_Z \qquad (C1)$$

and, using eq. (57), we get

$$\dot{\rho}_M = -3H\rho_M - \frac{\Omega_Z \rho_0}{H_0^2} H[2\dot{H} + 3(1 + w_Z)H^2]. \quad (C2)$$

The presence of the term proporional to $(1 + w_Z)$ on the right-hand side makes it more difficult to find an exact solution. It is however easy to find the solution perturbatively in Ω_Z, which is sufficient for our purposes since we know, from the successes of ΛCDM cosmology, that $\Omega_Z \ll 1$. Then, we search for a solution of the form

$$\rho_M(t) = \frac{1}{a^3}[\rho_M(t_0) + \Delta\rho_M(t)], \qquad (C3)$$

where, by definition, $\Delta\rho_M(t_0) = 0$ (we set as usual $a(t_0) = 1$), and we get

$$\frac{d}{dt}\Delta\rho_M = -\frac{\Omega_Z \rho_0}{H_0^2} a^3 H[2\dot{H} + 3(1 + w_Z)H^2]. \quad (C4)$$

We solve this equation perturbatively in Ω_Z, so to first order we simply replace the right-hand side of eq. (C4) by its value on the unperturbed solution $a(t) = (t/t_0)^{2/3}$ (assuming for simplicity a purely MD phase), with t_0 related to H_0 by $t_0 = 2/(3H_0\Omega_M^{1/2})$, and we get

$$\Delta\rho_M(t) = -2w_Z\Omega_Z\Omega_M\rho_0 \log\frac{t}{t_0}, \qquad (C5)$$

or, in terms of the redshift z,

$$\Delta\rho_M(z) = 3w_Z\Omega_Z\Omega_M\rho_0 \log(1 + z). \qquad (C6)$$

Therefore

$$\rho_M(z) = \rho_M(0)(1 + z)^3 [1 + 3w_Z\Omega_Z \log(1 + z)]. \quad (C7)$$

This expression is valid to first order in Ω_Z and, at this order, it is equivalent to

$$\rho_M(z) = \rho_M(0)(1 + z)^{3(1+w_Z\Omega_Z)}, \qquad (C8)$$

which for $w_Z = -1$ agrees with the exact result (68). Since the limits on Ω_Z discussed in Section IV basically come from the modified evolution of ρ_M with red-shift, we see that the limits on Ω_Z for $w_Z \neq -1$ can be obtained by replacing $\Omega_Z \to -w_Z\Omega_Z$ in the results of Section IV.

[1] S. Weinberg, Rev. Mod. Phys. **61** (1989) 1.
[2] P. J. E. Peebles and B. Ratra, Rev. Mod. Phys. **75**, 559 (2003).
[3] T. Padmanabhan, Phys. Rept. **380** (2003) 235.
[4] E. J. Copeland, M. Sami and S. Tsujikawa, Int. J. Mod. Phys. D **15** (2006) 1753.
[5] M. Maggiore, "A Modern Introduction to Quantum Field Theory," *Oxford University Press, 2005*, Section 5.7.
[6] R. L. Arnowitt, S. Deser and C. W. Misner (1962). In *Gravitation: an introduction to current research*, L. Witten ed., Wiley, New York, [arXiv:gr-qc/0405109].
[7] E. Poisson, "A Relativist's Toolkit. The Mathematics of Black-Hole Mechanics", *Cambridge University Press, 2004*.
[8] L. Parker and S. A. Fulling, Phys. Rev. D **9** (1974) 341.
[9] S. A. Fulling, L. Parker, Annals Phys. **87** (1974) 176.
[10] T. Padmanabhan, Class. Quant. Grav. **22** (2005) L107.
[11] E. K. Akhmedov, arXiv:hep-th/0204048.
[12] A. O. Barvinsky and G. A. Vilkovisky, Phys. Rept. **119** (1985) 1.
[13] I. L. Buchbinder, S. D. Odintsov and I. L. Shapiro, "Effective action in quantum gravity," *Bristol, UK: IOP (1992)*.
[14] I. L. Shapiro, Class. Quant. Grav. **25** (2008) 103001.
[15] A. M. Pelinson and I. L. Shapiro, arXiv:1005.1313 [hep-th].
[16] N. Bilic, arXiv:1004.4984 [hep-th].
[17] S. Basilakos, M. Plionis and J. Solà, Phys. Rev. D **80** (2009) 083511.
[18] E. Komatsu *et al.* [WMAP Collaboration], Astrophys. J. Suppl. **192**, 18 (2011).
[19] R. Amanullah *et al.*, Astrophys. J. **716** (2010) 712.
[20] I. Antoniadis, arXiv:hep-th/9909212.
[21] N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, Phys. Rev. D **59** (1999) 086004.
[22] I. L. Shapiro and J. Solà, JHEP **0202** (2002) 006.

[23] I. L. Shapiro, J. Solà, C. Espana-Bonet and P. Ruiz-Lapuente, Phys. Lett. B **574** (2003) 149.

[24] C. Espana-Bonet, P. Ruiz-Lapuente, I. L. Shapiro and J. Solà, JCAP **0402** (2004) 006.

[25] I. L. Shapiro and J. Solà, Phys. Lett. B **682** (2009) 105.

[26] J. Solà, J. Phys. A **A41** (2008) 164066.

[27] L. P. Grishchuk, Sov. Phys. JETP **40** (1975) 409.

[28] A. A. Starobinski, JETP Lett. **30** (1979) 682.

[29] A. A. Starobinsky, in *Field Theory, Quantum Gravity and Strings*, eds. H. J. de Vega and N. Sanchez, Springer Verlag (1986).

[30] J. Grande, J. Solà and H. Stefancic, JCAP **0608** (2006) 011.

[31] J. Grande, J. Solà and H. Stefancic, Phys. Lett. B **645**, 236 (2007).

[32] J. Grande, A. Pelinson and J. Solà, Phys. Rev. D **79** (2009) 043006.

[33] F. Bauer, J. Solà and H. Stefancic, JCAP **1012** (2010) 029.

[34] G. Mangano, G. Miele, S. Pastor, T. Pinto, O. Pisanti and P. D. Serpico, Nucl. Phys. B **729** (2005) 221.

[35] F. Iocco, G. Mangano, G. Miele, O. Pisanti and P. D. Serpico, Phys. Rept. **472** (2009) 1.

[36] M. Kowalski *et al.* [Supernova Cosmology Project Collaboration], Astrophys. J. **686** (2008) 749.

[37] A. G. Cohen, D. B. Kaplan and A. E. Nelson, Phys. Rev. Lett. **82** (1999) 4971.

[38] S. D. H. Hsu, Phys. Lett. B **594** (2004) 13.

[39] R. Horvat, Phys. Rev. D **70** (2004) 087301.

[40] R. Schutzhold, Phys. Rev. Lett. **89** (2002) 081302.

[41] More precisely, renormalization trades the cutoff Λ_c for the subtraction point μ. While the latter is in principle arbitrary, a clever choice of μ will minimize the effect of radiative corrections. For instance, if one renormalizes the electroweak theory in the MS scheme, one naturally takes $\mu = m_W$. Using a much lower subtraction point, say $\mu = m_e$, is in principle legitimate, but all radiative corrections would then become large, and the whole perturbative approach could be spoiled. In this sense, $\mu = m_W$ is a natural subtraction point for the electroweak theory, and we similarly expect that, in a theory involving quantum gravity, the natural subtraction point will be given by the Planck or string mass. (I thank a referee for this comment).

[42] Observe that in ref. [16] the result is written directly for the *complex* scalar belonging to a chiral superfield, hence the bosonic vacuum energy in eq. (32) of ref. [16] is twice as large as that in our eq. (22), which holds for a *real* scalar field.

[43] It is curious to observe that in all three case (De Sitter, RD and MD) the bare quantities $p_{\rm bare}$ and $\rho_{\rm bare}$ satisfy $p_{\rm bare} = w_{\rm bare}\rho_{\rm bare}$ with $w_{\rm bare} = (2/3)+w_{\rm dom}$ where $w_{\rm dom}$ is the w-parameter of the component that dominates the evolution during the corresponding phase, i.e. $w_{\rm dom} = -1, +1/3, 0$ during De Sitter, RD and MD, respectively.

[44] Observe that the value $1/(12\pi)$ of their numerical coefficients depends on the precise definition of the mass scale M, which in the approach based on RG involves not only the UV scale, but also some unknown beta-function coefficients, see eqs. (3.6) and (4.2) of ref. [24].

[45] Unless w_Z evolves in time so to track the equation of state of the dominant energy component, i.e. it evolves from $w_Z = 1/3$ during RD to $w_Z = 0$ during MD. Another interesting possibility, that we do not consider here, is that the Bianchi identities are actually satisfied by assuming a standard conservation law for matter, but assigning a time dependence to Newton's constant, see [26].

[46] The analysis was actually restricted to positive values of Ω_Z (J. Solà, personal communication).